Responsible Metrics and Scientific Evaluation Moving Towards Post-Print

Authors: Song Liping, Wang Xiuhuan, Yuan Shanshan

Date: 2023-04-01T16:02:50+00:00

Abstract

[Purpose/Significance] Taking the three foundational frameworks of responsible metrics—the San Francisco Declaration on Research Assessment (DORA), the Leiden Manifesto, and The Metric Tide report—as the starting point, this study interprets the development trends of scientific evaluation from the perspective of responsible metrics. [Method/Process] Based on Web of Science, this study visualizes papers and core authors from 2012 to 2020 using VOSviewer to reflect the diachronic changes in research hotspots within this field. [Results/Conclusion] Timeline-based keyword and author overlay visualizations indicate that a metrical transformation is underway, exhibiting a development trend centered on research quality, with researchers as the foundation, disciplinary normalization as the orientation, impact as the radius, citation exploration as the basis, and altmetrics as the reference. The conclusion points out that governing scientific evaluation through responsible metrics is the essential path to establishing a healthy academic ecosystem, and that the implementation of responsible metrics signifies a transformation of deeply entrenched incentive mechanisms and metric cultures.

Full Text

Responsible Metrics and the Trend of Scientific Evaluation

Song Liping, Wang Xiuhang, Yuan Shanshan

School of Management, Tianjin Normal University, Tianjin 300387

Abstract: [Purpose/Significance] This paper interprets the development trend of scientific evaluation from the perspective of responsible metrics, based on its three foundational frameworks: the San Francisco Declaration on Research Assessment, the Leiden Manifesto, and the Metric Tide report. [Method/Process] Using Web of Science and VOSviewer, we visualize papers and core authors from 2012-2020 to reflect diachronic changes in research hotspots in this field.

[Result/Conclusion] Timeline-based keyword and author overlay views demonstrate that a revolution in research metrics is underway, presenting a development trend centered on research quality, with researchers as the foundation, discipline normalization as the orientation, impact as the radius, citation exploration as the root, and altmetrics as the reference. The conclusion points out that governing scientific evaluation through responsible metrics is the only path to establishing a healthy academic ecosystem, and that implementing responsible metrics implies a transformation of deeply entrenched incentive mechanisms and metric cultures.

Keywords: Responsible metrics; Scientific evaluation; San Francisco Declaration; Leiden Manifesto; Metric Tide

Funding: This paper is a result of the National Social Science Fund General Project "Research on Optimization of Scientific Evaluation Methods and Indicators from the Perspective of Responsible Metrics" (Project No.: 18BTQ074).

Author Biographies: - **Song Liping** (ORCID: 0000-0003-3056-3720), Professor, Ph.D., Master's Supervisor - **Wang Xiuhang** (ORCID: 0000-0003-1974-5080), Master's Student, Corresponding Author, E-mail: wzxh118@163.com - **Yuan Shanshan** (ORCID: 0000-0002-2736-3333), Master's Student

Received: December 24, 2020; **Revised**: February 9, 2021

The world of research evaluation is undergoing transformation. In 2015, Professor J. Wilsdon et al. from the University of Sheffield proposed the concept of "responsible metrics" in *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (hereinafter referred to as the "Metric Tide report") [**?**]. This report, together with the *San Francisco Declaration on Research Assessment* (hereinafter referred to as the "San Francisco Declaration") and the *Leiden Manifesto*, constitutes the main framework of responsible metrics. In 2018, J. Wilsdon further wrote in *The Guardian* that the metric tide is surging toward responsible metrics [**?**]. This paper analyzes the development trend of scientific evaluation from the perspective of responsible metrics, starting from these three programmatic documents.

1. Background Review

The San Francisco Declaration emerged from the 75th annual meeting of the American Society for Cell Biology held in San Francisco on December 16, 2012. It advocates developing robust and efficient assessment methods that do not rely on journal impact factors, reflecting the scientific community's strong call to correct the practice of evaluating papers based on journals and to change how research output quality is evaluated [**?**].

The Leiden Manifesto originated from the seven principles for the rational use of scientific evaluation indicators proposed by Professor D. Hicks from Georgia Institute of Technology at the International Conference on Science and Technology Indicators (STI) held in Leiden, Netherlands, in 2014. In April 2015,

D. Hicks, together with L. Waltman, Director of the Centre for Science and Technology Studies at Leiden University, and P. Wouters, expanded these into ten principles and published them in *Nature*. These principles distill the norms of indicator-based research evaluation and herald the brewing transformation in research assessment [**?**].

The Metric Tide report was a product of the UK's Research Excellence Framework (REF), a metrics-based higher education evaluation system first implemented in 2014. Drawing on the EU's "Responsible Research and Innovation" governance framework, it proposed approaches for the rational use of quantitative indicators in research evaluation, defining responsible metrics as robust, diverse, multi-dimensional, transparent, and reflexive practices based on accurate data that reflect disciplinary differences. This interpretation established the conceptual framework of responsible metrics [**?**].

These three programmatic documents represent new thinking in scientific evaluation, constructing the intellectual space for responsible metrics and establishing new norms for scientific assessment. Under their guidance, many countries have followed responsible metrics principles in practice. Bibliometricians such as P. Wouters, L. Waltman, L. Bornmann, and M. Thelwall have integrated responsible metrics principles into explorations of altmetrics applications, research impact, and other issues [**?**]. At the 23rd STI conference in 2018, L. Waltman delivered a keynote speech on "Responsible Metrics," and that same year, the UK Forum on Responsible Metrics released a survey report titled "Changing the Metric Culture" [**?**]. These developments demonstrate the changes that responsible metrics has brought to scientific evaluation theory and practice.

2. Visualization of Research Hotspots from 2012-2020

To present the research hotspots and development trends in responsible metrics in recent years, this study uses Web of Science (WoS) and VOSviewer, a visualization tool developed by N.J. van Eck and L. Waltman at Leiden University based on JAVA, to visualize papers and core authors from 2012-2020 [**?**]. The core idea of this software is to reveal research hotspots, author groups, and other patterns through co-occurrence clustering and visual analysis, showcasing the structure, evolution, and collaboration relationships within knowledge domains. The search was conducted on June 26, 2020, using the query: TS="responsible metric*" OR TS="research evaluat*" OR TS="research assessment*" OR TS="science policy*" OR TS="impact factor*" OR TS="journal impact factor*" OR TS="citation impact*" OR TS="altmetric*" OR TS="article level metric*" OR TS="scholarly impact*" OR TS="academic impact*" OR TS="citespace*" AND Publication years=2012-2020 and Web of Science categories=information science library science, retrieving 1,778 papers. We then selected timeline-based keyword and author overlay views for visualization to reflect the diachronic changes and development trends of research hotspots in this field.

As clearly shown in Figure 1 [Figure 1: see original paper] and Figure 2 [Figure 2: see original paper], the evolution path begins with explorations of journal impact factors and webometrics, gradually transitioning to research evaluation, science policy, research quality, and researchers, and more recently extending to citation impact (including field-normalized citation impact), altmetrics (applications), and studies on collaboration and innovation. The author overlay view based on timeline indicates that L. Waltman, P. Ingwersen, R. Rousseau, L. Leydesdorff, L. Bornmann, M. Thelwall, H.F. Moed, and W. Glanzel have become core figures, while V. Larivière, K. Kousha, and R. Haunschild have emerged as recent active contributors. How should we interpret these results?

3. Interpretation of Scientific Evaluation Development Trends

Figure 1 outlines the development trajectory of scientific evaluation under the responsible metrics framework, revealing characteristics centered on research (including research quality, researchers, and disciplinary differences) with impact (including citation impact, societal impact, and altmetrics applications) as the radius. As a significant event in the history of scientometrics, this trajectory not only marks an ongoing transformation but also indicates future trends in scientific evaluation. The following sections provide specific analysis based on recent responsible metrics practices.

3.1 Research Quality as the Center

The "research quality" cluster formed in Figure 1 and recent research evaluation practices demonstrate the trend of responsible metrics centering on research. Fundamentally, as an extension of the EU's "Responsible Research and Innovation" concept in research governance to the field of scientific evaluation, responsible metrics has been inseparable from scientific research since its inception. The San Francisco Declaration aims to establish "research-centered" assessment methods, while the Leiden Manifesto warns that "we must have full awareness that evaluation indicators change researchers' motivations and thus the entire research system" [**?**], reminding us to treat the relationship between evaluation and research with caution.

However, since the intensification of publisher competition in the latter half of the 20th century, the original purpose of impact factors in journal selection has been distorted, leading journals to add a fifth function—evaluation—on top of their traditional roles of registration, certification, dissemination, and preservation. Impact factors gradually evolved into labels for research paper quality. The simplistic practices of "evaluating papers by journals," "quantification supremacy," and crude "piece-rate" evaluation systems deviated from the goals of scientific evaluation, ignoring the complex interactions between evaluation and knowledge production. Consequently, the academic ecosystem elements—researchers, scientific research, and journals—became nested in an evaluation

system based on journal impact factors. Papers became the "hard currency" of academia, and impact factors became an "axiom requiring no proof" for "paper quality" [**?**]. Some funding agencies in the Czech Republic and Italy require applicants to list impact factors alongside their publications; Belgium's formula for allocating funds to universities includes weighting based on impact factors; Brazil's Ministry of Education system Qualis uses impact factors to determine funding allocation [**?**]. This neglect of research quality has forced researchers to sacrifice innovative research and pick "low-hanging fruit" instead of proposing bold hypotheses, seeking safe publication in high-impact journals. This has led to the prevalence of "CNS worship" (worship of *Cell*, *Nature*, *Science*), where publishing in authoritative journals becomes the main publication strategy, causing scientific research to seriously deviate from Mertonian scientific norms, with the scientific enterprise ultimately paying the price. As J. Lane's famous quote reminds us: "Flawed metrics produce flawed science" [**?**]. One cannot help but question: what is the point of evaluating research quality by the publication venue?

Therefore, S. Curry, chair of the San Francisco Declaration steering committee and a structural biologist at Imperial College London, stated in an interview with *Nature* that "any use of impact factors should carry a clear warning label, like cigarettes, as they are the root of many evils in current academic evaluation systems" [**?**]. Responsible metrics advocates returning to the true host of research quality: research itself. Moreover, at the annual meeting of the UK Bibliometric Society in March 2020, Professor S. Hill, chair of the REF2021 steering committee and Director of Research at Research England, included research and research culture as two of the four drivers shaping the future of research evaluation in his speech titled "Four Trends Shaping the Future of Research Evaluation," signaling that the trend of centering on research and research quality will continue [**?**].

3.2 Researchers as the Foundation

Within the research-centered domain, the emergence of a "researchers" cluster in early 2017 in Figure 1 indicates responsible metrics' attention to researchers (including female researchers). L. Bornmann once profoundly revealed researchers' mimetic behavior toward scientific evaluation indicators through the concept of "mimesis in science." The Metric Tide report cites the tragedy of an Imperial College professor: "flawed evaluation criteria distort our behavior, determine our career paths, and even lead to what the Archbishop of Canterbury calls the new barbarism in universities" [**?**]. Objectively speaking, in an impact factor-based evaluation system, publishers, funders, research institutions, and researchers are all "prisoners of dilemma." However, publishers dominate by bundling with impact factors, funders by designing game rules, while researchers can only succumb to the constraints of established evaluation indicators. "Indicator dependency" under crude research management means that publishing trumps scientific discovery, forcing researchers to suppress their curiosity—the original

driving force of scientific research—and maximize research output metrics to meet indicators. Statistics show that German researchers have changed their publication strategies to publish in high-impact English journals to increase citations; Australia has seen a steady rise in SCI papers while experiencing a noticeable decline in citations; A.B. Wagner from the University at Buffalo even proposed "Get Cited or Perish" as an upgraded version of "Publish or Perish," making it researchers' motto. Survival pressure, "obsession" with impact factors, and responses to improper incentives have fostered opportunism in academia, leading to salami publishing, honorary authorship, citation cartels, and research fraud. Goodhart's Law in economics provides the best explanation for this phenomenon: when a metric becomes a target, it ceases to be a good metric, because people start gaming it [**?**].

Given the coupling effects among scientific evaluation, the scientific community, science policy, and scientific research, responsible metrics as a paradigm of scientific evaluation strengthens dialogue with researchers and incorporates them as a reference system into the research evaluation system. In 2017, Clarivate Analytics and the Centre for Science and Technology Studies at Leiden University convened researchers for a symposium titled "Rethinking Impact Factors: New Approaches to Journal Evaluation." In 2019, RAND Europe, based on a survey of 3,600 researchers, formed a report titled "The Changing Research Landscape and Reflections on National Research Assessment" for REF2021 reference. The Howard Hughes Medical Institute gathered researchers from over 30 institutions in North America and Europe to jointly explore the implementation of the San Francisco Declaration. Today, the three programmatic documents, originally "soft constraints" on scientific evaluation by the scientific community, are gradually becoming hard policies in the EU, Australia, the UK, and other countries. This operation signifies exploration and reflection on metric culture and incentive mechanisms, revealing the beginnings of a people-oriented responsible metrics culture.

3.3 Discipline Normalization as the Orientation

Along with the deepening of responsible metrics practices and the further diffusion of the research-centered trend, a discipline normalization cluster appears in Figure 1. In the WoS literature retrieval list, "field-normalization" also emerges as a high-frequency term, reflecting the attention to responsible metrics theory and practice. Fundamentally, the "heat" around academic impact normalization in responsible metrics originates from the need to compare different disciplines in the UK's REF 2014 implementation. Due to the skewed nature of citations—a characteristic where citation density varies across disciplines, publication years, and document types—E. Garfield pointed out in 1979 that raw citation frequencies across different disciplines cannot be directly compared, making discipline normalization necessary [**?**]. As one normalization method, field normalization aims to map values with different ranges onto the same scale, either [0,1] or [-1,1], to eliminate disciplinary noise in scientometrics. The emergence of a

recent active research group centered on R. Haunschild in the author overlay view of Figure 2 further indicates the exploration of discipline normalization algorithms in responsible metrics practice. In 2016, R. Haunschild from the National Institutes of Health proposed RCR (Relative Citation Ratio), which dynamically defines a paper's discipline based on co-citation networks and normalizes citation indicators, thereby avoiding constraints from hierarchical structures and dynamic development of disciplines on indicators like CNCI (Category Normalized Citation Impact) and providing new solutions for evaluating and comparing citation impact of interdisciplinary, particularly cross-disciplinary, research. Undoubtedly, cross-disciplinary research has significant advantages in promoting teamwork, solving challenging problems, integrating fragmented academic worlds, and reshaping academic landscapes. However, the current paradox is that while governments encourage cross-disciplinary research, its value often receives delayed recognition or is even ignored due to unconventional evaluation methods [**?**]. In 2016, Lord Stern, President of the British Academy and Chief Economist of the World Bank, emphasized in the Stern Review for REF2014 that "interdisciplinary research plays an important role in solving complex global social, economic, ecological, and political problems," while also expressing concern that interdisciplinary research receives unfair treatment in peer review [**?**]. Although Cambridge University Professor D.A. Donald, who conducts interdisciplinary research in physics and biology, formed the REF2021 Interdisciplinary Research Assessment Panel and promised impartial evaluation policies [**?**], REF2021 submissions of relevant case applications remain far below expectations, all of which seem to question the scientific evaluation system. Meanwhile, the formation of cooperation and innovation clusters in Figure 1 in 2017, explorations of CNCI and other field-normalized citation impact based on a priori classifications, and the development of indicators like RCR that identify research fields through co-citation clustering further reflect how to understand and support diverse and dynamic academic research ecosystems, particularly cross-disciplinary research, has become a focal point of responsible metrics.

3.4 Impact as the Radius

Figure 1 shows that scientific evaluation under the responsible metrics wave, while spreading outward from research as the core, has impact as its keyword. Among the 1,778 retrieved documents, 415 titles contain the word "impact," and L. Waltman's "A review of the literature on citation impact indicators" was cited 284 times in the same publication year, confirming this wave's attention to research impact. In fact, based on cost-benefit theory in economics, whether REF2014 or REF2021, impact is the main theme. REF2014 first centered its assessment on impact, defining it as the influence and benefits on economy, society, culture, public policy, services, health, environment, or quality of life beyond academia. The UK Science and Technology Minister and Chancellor of the Exchequer specially commissioned Lord Stern to conduct the Stern Review for REF2014, and REF2021 implementation details further increased the weight of impact [**?**]. RAND Europe's notion of "maximizing the benefits that

science, technology, and innovation bring to society and economy by strengthening impact" reveals the origin of this approach [**?**]. Specifically, responsible metrics-based scientific evaluation has been regarded as a key means to enhance UK's research capacity and quality and a lever to promote scientific and technological progress, with the path to achieving this goal being to prioritize impact, betting on impact-based resource allocation and incentives for institutional and individual behavior.

Academic impact and societal impact represent two dimensions of impact evaluation. Based on R.K. Merton's normative theory—that "citations are the currency used to repay research debts and represent influence on scientific research work or cognition"—academic impact is quantified as a "dependent variable" of citations [**?**]. Meanwhile, with the development of the internet, impact has diffused to societal impact, called a "Kuhnian revolution" in evaluation [**?**]. To this end, responsible metrics seeks "citation equivalents" to measure societal impact. King's College London and Digital Science conducted data mining on 7,000 impact cases in REF2014, extracting common quantitative indicators of societal impact. In 2018, RAND Europe, commissioned by the Higher Education Funding Council for England (HEFCE), released REF case analysis impact quantitative indicator standards and developed Impact Finder, an analytical suggestion package consisting of 300 questions to collect research impact on society, culture, and economy. L. Bornmann and R. Haunschild used convergent validity to compare the similarity between case studies and altmetrics in measuring societal impact results [**?**]. Research shows that currently, aside from patent technology impact and medical clinical impact as two citation-equivalent societal impact measurement methods, case studies are the only feasible approach. Given that citations are already an academic consensus as a quantifiable indicator of academic impact, the measurement of societal impact as a new quantitative element in this responsible metrics wave has become an issue under exploration.

3.5 Citation Exploration as the Criterion

Regarding citation impact, Figure 1, Figure 2, UK REF practices, the Metric Tide report, and S. Hill's assertions [**?**] all indicate that citations remain the foothold of scientific evaluation, whether in the past, present, or future. Specifically, under the impact of the responsible metrics wave, citations remain the criterion for scientific evaluation. In fact, citations serve as the standard weight, and the equivalence between citations and academic impact remains the major premise of scientific evaluation. At its root, first, science is a gift economy whose value is expressed through its contribution to the total body of knowledge or its influence on others. Since authors use citations to indicate influence from other researchers and their work, academic impact and citations have relevance. Second, the essence of citation is knowledge flow—what Thomson Reuters Chief Scientist H.H. Small called "citations as signposts left after information is used" [**?**]—showing the complex process of knowledge unit dispersion and recombi-

nation, inheritance and innovation, evolution and sublimation. Thus, using citations as traces of scientific communication and impact as identifiers allows us to trace scientific development patterns, optimize scientific communication, and advance scientific research. Third, citation analysis can identify and map research fronts, define disciplines and emerging specialties, and determine the interdisciplinary characteristics and impact of research programs and projects. Therefore, citation and citation analysis have high compatibility with the overall goals of scientific evaluation. Consequently, even though S. Woolgar questioned citation analysis based on the assertion that "citations are a function of many variables besides scientific impact" [**?**], and M.H. MacRoberts accused academia of having a conservative "Cassandra complex" in refusing to abandon citation analysis [**?**], citations' position in scientific evaluation has never been shaken.

It should be noted that scientific evaluation under responsible metrics is conducted against the backdrop of Citation Analysis 4.0. In addition to explorations of paper-level impact issues such as RCR research and CNCI, full-text bibliometric analysis has become the focus, where identifying intersections between citing and cited documents enables a deeper understanding of citations and motivations in scientific papers, more accurately reconstructs scientific communication processes, and ultimately better serves scientific evaluation. This represents a development trend for the foreseeable future.

3.6 Altmetrics as the Reference

Figure 1's altmetrics, social media metrics, and Elsevier's ICSR (International Center for the Study of Research) diversified scientific evaluation plans demonstrate that altmetrics holds a place in scientific evaluation. Since J. Priem proposed altmetrics based on online scientific communication in 2010, AM values have entered the field of vision as a supplement to traditional bibliometrics, triggering reflection on measurement methods and catalyzing multi-dimensional evaluation at the paper level [**?**]. In *Nature*'s 2013 impact exploration album, the scientific community reached consensus on the multi-dimensionality of impact. The Stern Review also called for a broader definition of research impact. Meanwhile, the National Information Standards Organization (NISO) established altmetrics usage standards. L. Bornmann and R. Haunschild noted that "the ten principles of the Leiden Manifesto have important guiding significance for altmetrics" [**?**]. J.D. Jenkins defined responsible metrics as a combination of rationally used citation indicators, altmetrics indicators, and other quantitative assessment methods, thus setting the tone for altmetrics as a reference system for citations in responsible evaluation [**?**].

However, although some experts believe altmetrics has brought revolutionary changes to academic research and its impact measurement, M. Thelwall [**?**] and H.F. Moed [**?**] have pointed out that only Mendeley reader statistics and F1000 peer review results have relatively high correlation with citation statistics and can be used for early prediction of impact; Twitter and other altmetrics indicators mainly reflect "attention" rather than impact. Therefore, the effectiveness

of altmetrics remains an unresolved issue. On the one hand, bibliometricians test its convergence with existing indicators through correlation analysis to interpret its meaning; on the other hand, people transplant relatively mature bibliometric methods such as co-citation analysis and CNCI into altmetrics to improve its standards and establish diversified scientific evaluation methods oriented toward the entire workflow of scientific creation, dissemination, and utilization. Indeed, just as negative citations and excessive self-citations interfere with citation analysis, data quality and cleanliness seriously affect analysis results. Therefore, how to screen and validate altmetrics data sources, clean data, and organically combine contextual analysis and statistical correlation analysis with peer review and bibliometrics to interpret its connotation and improve its reliability and validity are key issues for future altmetrics application in responsible metrics.

Conclusion

Nowadays, scholars worldwide and EU Commission expert panels have identified research evaluation as the cornerstone for constructing a global research architecture. The impact factor-based evaluation system has been called "a cancer that can no longer be ignored" and even "numbers that devour science" [**?**], while China's science and technology policy is steadily moving away from impact factor-based reward measures. This shows that governing academic evaluation through responsible metrics is the only path to establishing a healthy academic ecosystem, and implementing responsible metrics implies a transformation of deeply entrenched incentive mechanisms and metric cultures.

Based on scientific evaluation practices in Europe—the birthplace of responsible metrics—and research hotspots reflected by VOSviewer, with the three main frameworks of the San Francisco Declaration, Leiden Manifesto, and Metric Tide report as the core, responsible metrics culture is forming and presents a development trend centered on research quality, with researchers as the foundation, discipline normalization as the orientation, impact as the radius, citation exploration as the root, and altmetrics as the reference. Although implementation has been tortuous due to the inertia of scientific evaluation and vested interests, Professor S. Curry has announced the San Francisco Declaration Implementation Roadmap; E. Gadd, chair of the International Society for Research Management's Research Evaluation Working Group, has developed the SCOPE framework with high sensitivity to evaluation purposes; the evaluative inquiry proposed by ICSR has strong operability; the Metrics Toolkit is already online; the CRediT taxonomy for academic contributions has been implemented; and artificial intelligence will help us more completely reconstruct the true face of evaluation objects without repeating the story of "the blind men and the elephant" due to one-sided understanding. All these confirm the implementation footsteps of responsible metrics. On the one hand, we have reason to believe that responsible metrics will not repeat the past forty years where people turned a deaf ear to various abuses after E. Garfield proposed preventing citation anal-

ysis misuse in 1977 [**?**]. On the other hand, we must 清醒地意识到 that an ideal responsible metrics with stability, multi-dimensionality, diversity, transparency, and iterability is still under exploration. As Professor J. Wilsdon pointed out in his 2020 lecture titled "New Frontiers in Research Evaluation": "We have now taken the first step in the right direction, but we still need to develop better indicators, embed principles with more effort, and establish more sensitive management frameworks." Meanwhile, it should be noted that compared with quantitative indicators, unstructured peer review will approach the true state of the evaluated object to the greatest extent. Therefore, in this uncertain world, the organic combination of responsible metrics-based heuristics and metrics-informed peer review will become the future direction of scientific evaluation [**?**].

References

[1] WILSDON J, ALLEN L, BELFIORE E, et al. The metric tide: report of the independent review of the role of metrics in research assessment and management[M]. London: SAGE Publications Ltd, 2015.

[2] WILSDON J. Has the tide turned towards responsible metrics in research?[N]. The guardian, 2018-07-10(1).

[3] San Francisco declaration on research assessment[EB/OL]. [2020-09-23]. http://www.ascb.org/dora/.

[4] HICKS D, WOUTERS P, WALTMAN L, et al. Bibliometrics: the Leiden manifesto for research metrics[J]. Nature, 2015, 520(7548): 429-431.

[5] CASSIDY R S. Theories of informetrics and scholarly communication[M]. Berlin: De Gruyter, 2016.

[6] WILSDON J. UK progress towards the use of metrics responsibly: three years on from The Metric Tide report[EB/OL]. [2021-03-23]. https://dera.ioe.ac.uk/31945/1/UK%20progress%20towards%20the%20use%20of%20metrics%20responsibly%

[7] Li Jie. Principles and applications of scientific knowledge mapping: a beginner's guide to VOSviewer and CitNetExplorer[M]. Beijing: Higher Education Press, 2018.

[8] GUÉDON J C, KRAMER B, LAAKSO M, et al. Future of scholarly publishing and scholarly communication: report of the expert group to the European commission[EB/OL]. [2020-09-06]. https://op.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1.

[9] STEPHAN P, VEUGELERS R, WANG J. Reviewers are blinkered by bibliometrics[J]. Nature, 2017, 544(7651): 411-412.

[10] LANE J. Let's make science metrics more scientific[J]. Nature, 2010, 464(7288): 488-489.

[11] CURRY S. Let's move beyond the rhetoric: it's time to change how we judge research[J]. Nature, 2018, 554(7691): 147.

[12] HILL S. Four trends shaping the future of research evaluation[EB/OL]. [2020-09-06]. https://figshare.com/articles/Four_{{{trends}}}{{shaping}}{{{the}}}{{future}}{{{of}}}{{resea

[13] WOUTERS P, SUGIMOTO C R, LARIVIÈRE V, et al. Rethinking impact factors: better ways to judge a journal[J]. Nature, 2019, 569(7758): 621-623.

[14] WALTMAN L. A review of the literature on citation impact indicators[J]. Journal of informetrics, 2016, 10(2): 365-391.

[15] VAN NOORDEN R. Interdisciplinary research by the numbers[J]. Nature, 2015, 525(7569): 306-307.

[16] STERN N. Lord Stern's review of the research excellence framework: call for evidence[EB/OL]. [2020-09-08]. https://core.ac.uk/download/pdf/74380480.pdf.

[17] Research excellence framework 2021[EB/OL]. [2020-09-08]. https://www.ref.ac.uk/.

[18] PARKS S S. The changing research landscape and reflections on national research assessment in the future[EB/OL]. [2020-09-08]. https://re.ukri.org/documents/2019/rand-europe-full-report/.

[19] BORNMANN L. Scientific revolution in scientometrics: the broadening of informetrics and scholarly communication[M]//CASSIDY R S. Theories of informetrics and scholarly communication. Berlin: De Gruyter, 2016: 347-359.

[20] BORNMANN L, HAUNSCHILD R. Do altmetrics assess societal impact in a comparable way to case studies? an empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF)[J]. Journal of informetrics, 2019, 13(1): 325-340.

[21] MACROBERTS M H, MACROBERTS B R. The mismeasure of science: citation analysis[J]. Journal of the Association for Information Science & Technology, 2017, 69(3): 474-482.

[22] BORNMANN L, HAUNSCHILD R. To what extent does the Leiden manifesto also apply to altmetrics? a discussion of the manifesto against the background of research into altmetrics[J]. Journal of the Association for Information Science & Technology, 2016, 40(4): 529-543.

[23] DAVID J. What are responsible metrics[EB/OL]. [2020-09-06]. http://www.open.ac.uk/blogs/the_{orb}/?p=3242.

[24] THELWALL M. Measuring societal impacts of research with altmetrics? common problems and mistakes[EB/OL]. [2020-09-08]. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/

[25] MOED H F. Applied evaluative informetrics[M]. Berlin: Springer International Publishing, 2017.

[26] LARIVIÈRE V, SUGIMOTO C R. The journal impact factor: a brief history, critique, and discussion of adverse effects[C]//GLÄNZEL W, MOED H

F. Springer handbook of science and technology indicators. Berlin: Springer, 2019: 3-24.

[27] BORNMANN L, HUG S, MAREWSKI J N. Bibliometrics-based heuristics: what is their definition and how can they be best studied?[EB/OL]. [2020-09-08]. https://arxiv.org/abs/1810.13005v1.

Author Contributions: - **Song Liping**: Topic selection and writing - **Wang Xiuhang**: Data analysis - **Yuan Shanshan**: Data collection

Note: Figure translations are in progress. See original paper for figures.